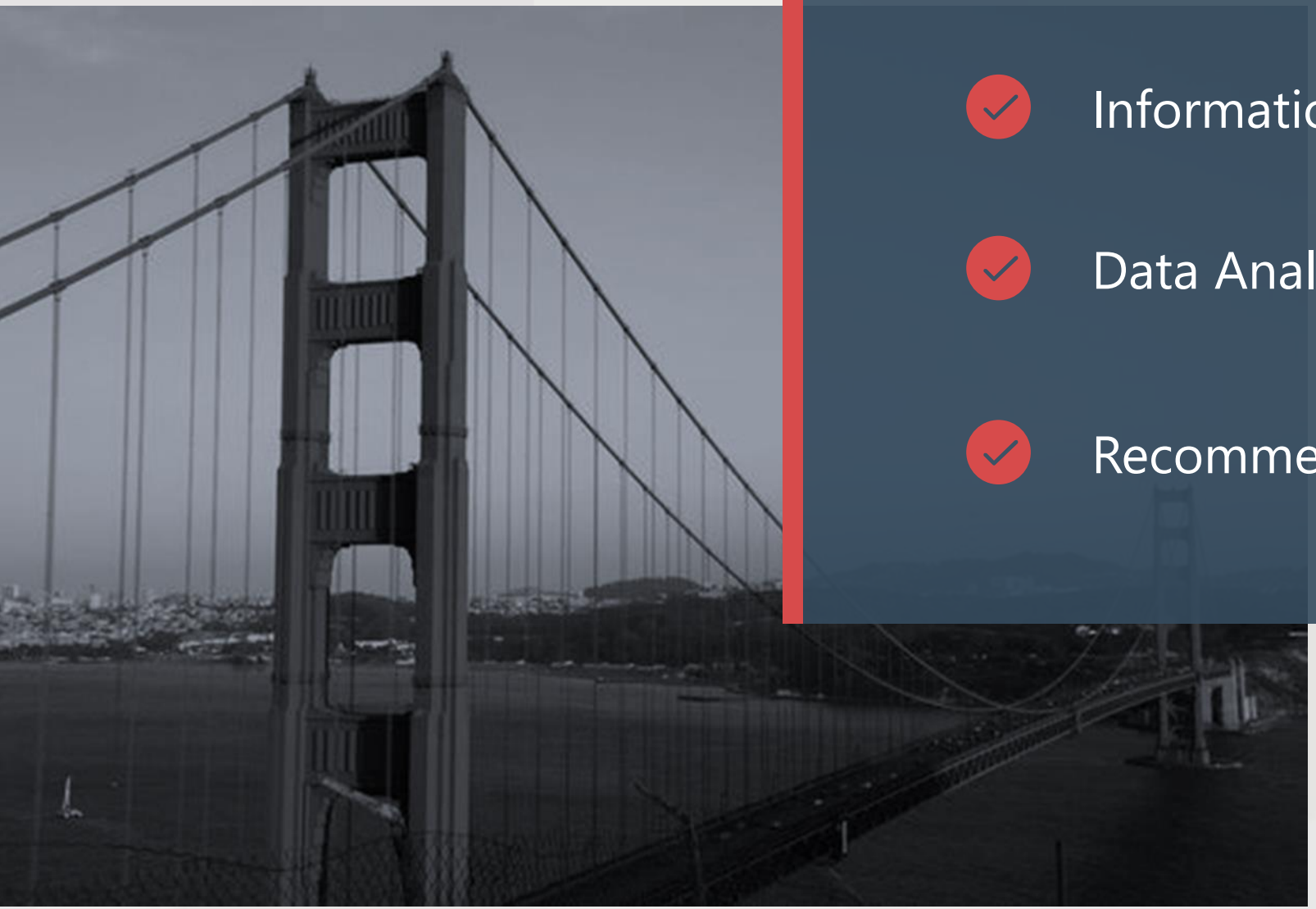
San Francisco Crime Analysis

GROUP 3



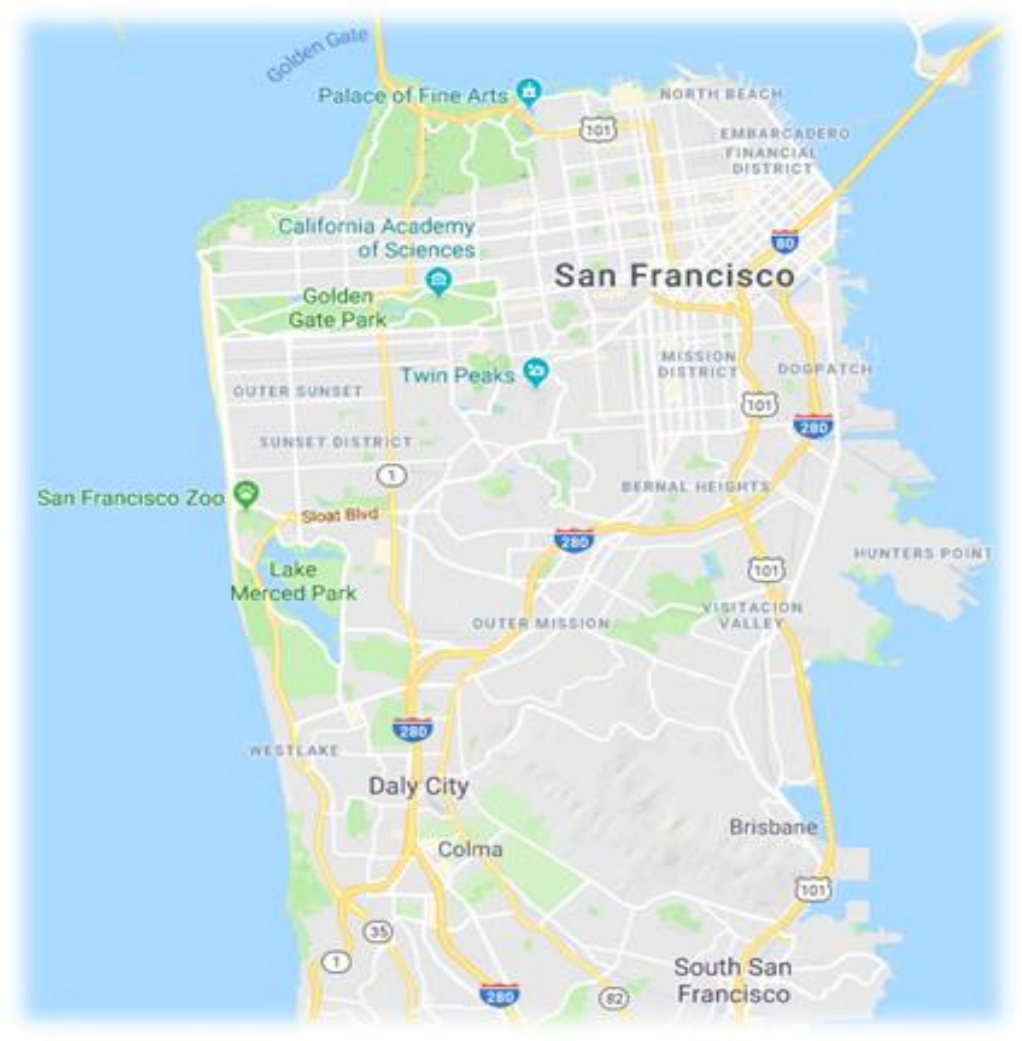
OUTLINE

- Information about San Francisco
- Data Analysis
- Recommendations

1

INTRODUCTION

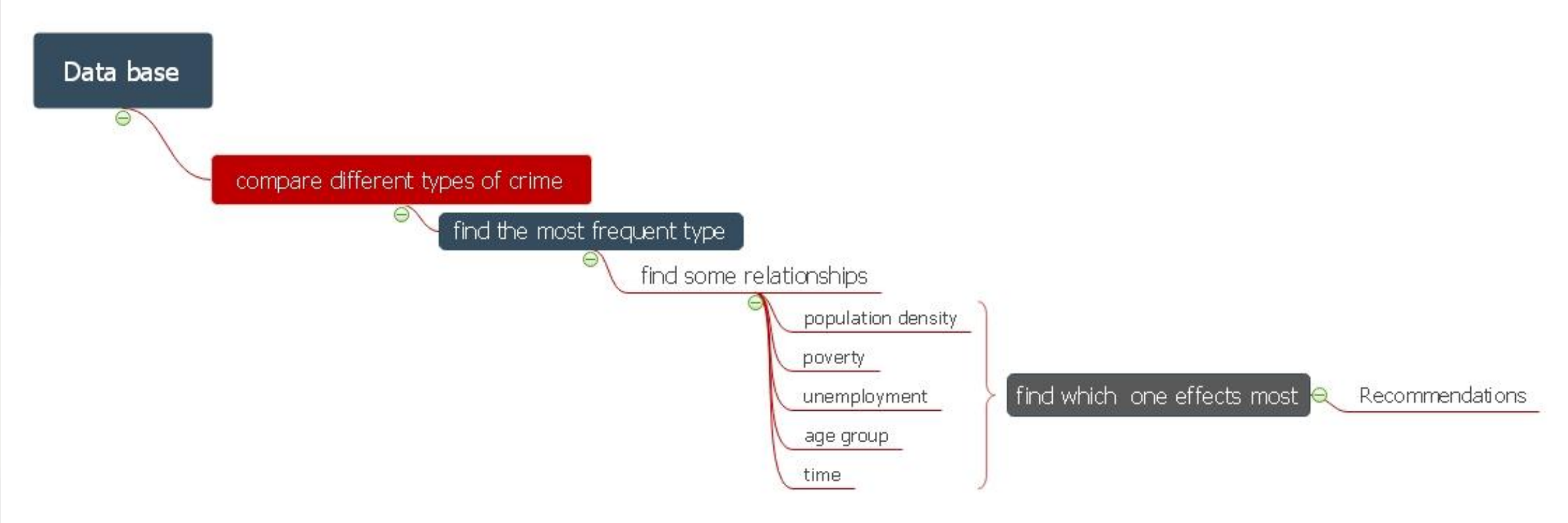
San Francisco

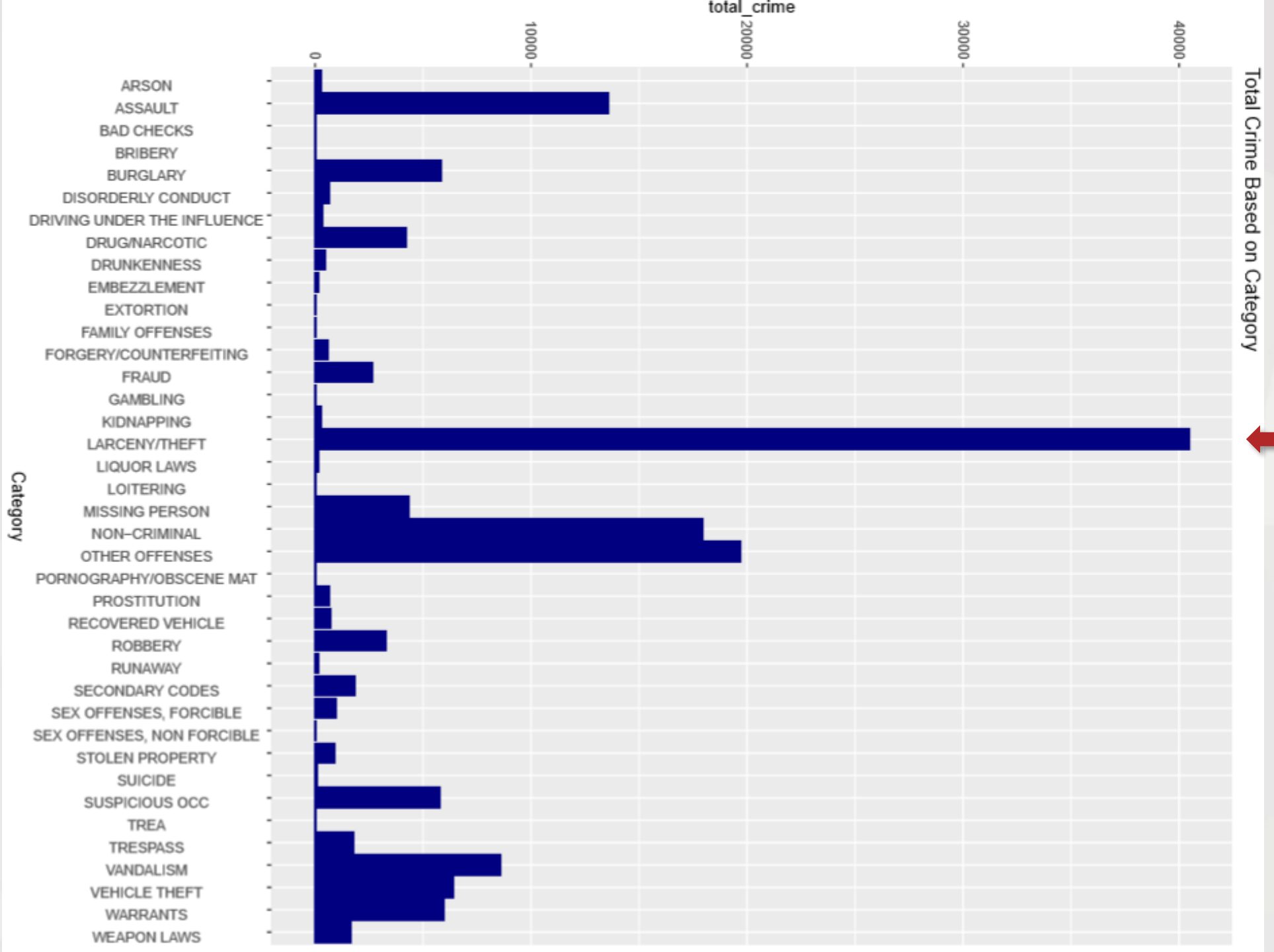


San Francisco officially called the City and County of San Francisco, is a city in, and the cultural, commercial, and financial center of, Northern California.

San Francisco is the 13th-most populous city in the United States, and the fourth-most populous in California, with 883,305 residents as of 2018.It covers an area of about 46.89 square miles (121.4 km2).

ABSTRACT


Data base
compare different types of crime
find the most frequent type
find some relationships
population density
poverty
unemployment
age group
time
find which one effects most
Recommendations



The most frequent type

2

DATA ANALYSIS

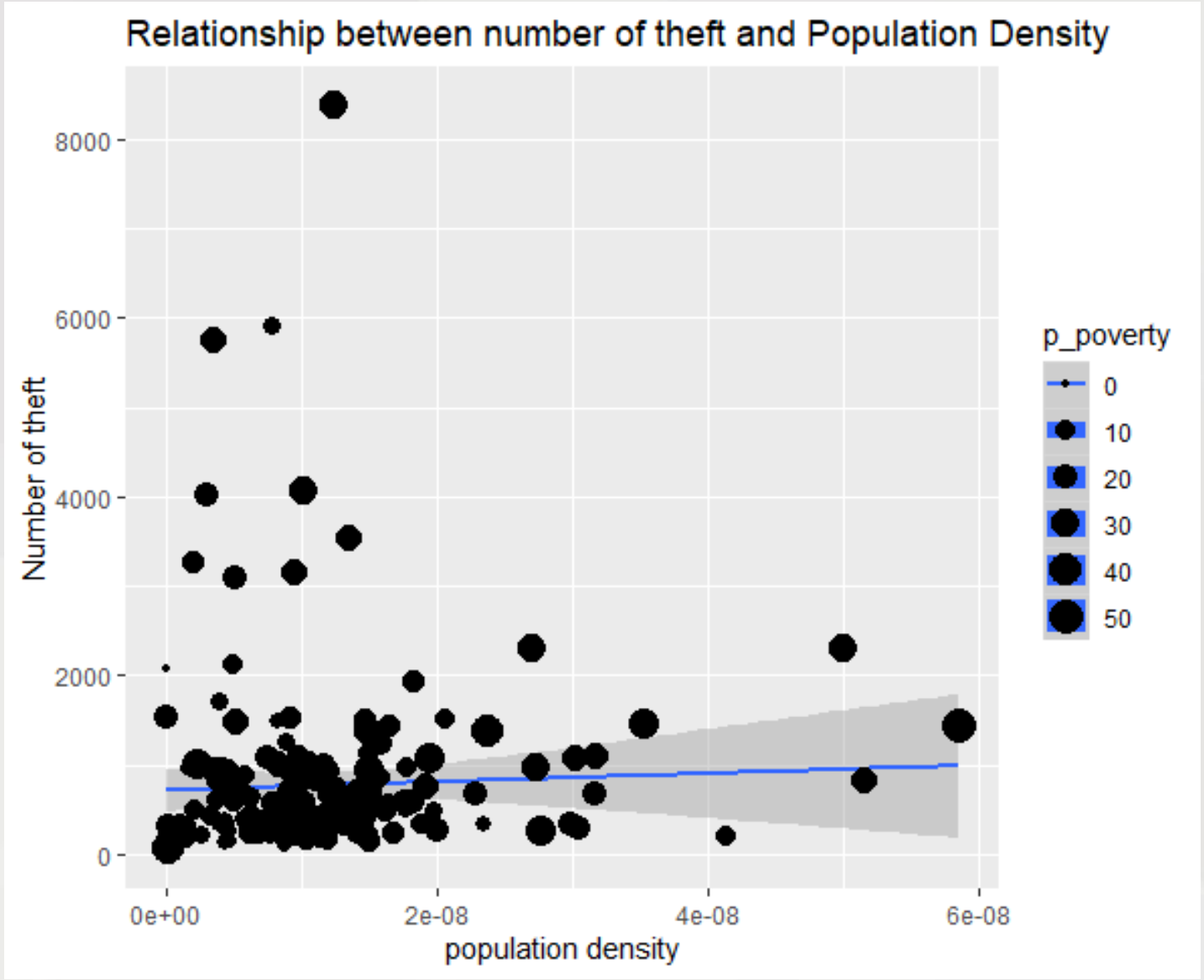
Population density

1.> *TO PERFORM INNER JOIN*

```
dsPlot <-   crime_SFO  %>%
filter ( Category == 'LARCENY/THEFT' )  %>%
group_by ( Tract )   %>%
summarise ( total_crime = n() )  %>%
inner_join ( census_SFO, by =  c ( "Tract" = "censustract") )
```

2.> *TO CREATE A SCATTER PLOT*

```
dsPlot  %>%
ggplot( aes ( x = pop_den,  y = total_crime, size =p_poverty ) ) +
geom_smooth ( method = 'lm' ) +
geom_point ( ) +
labs ( title ="relationship with population density",x="Population density",y="Number of theft/larceny")
```



- It seems that there is no significant relationship between the number of theft and population density,

- However, in when population becomes larger, the confident interval also increases, therefore, there is more uncertainty of our result.

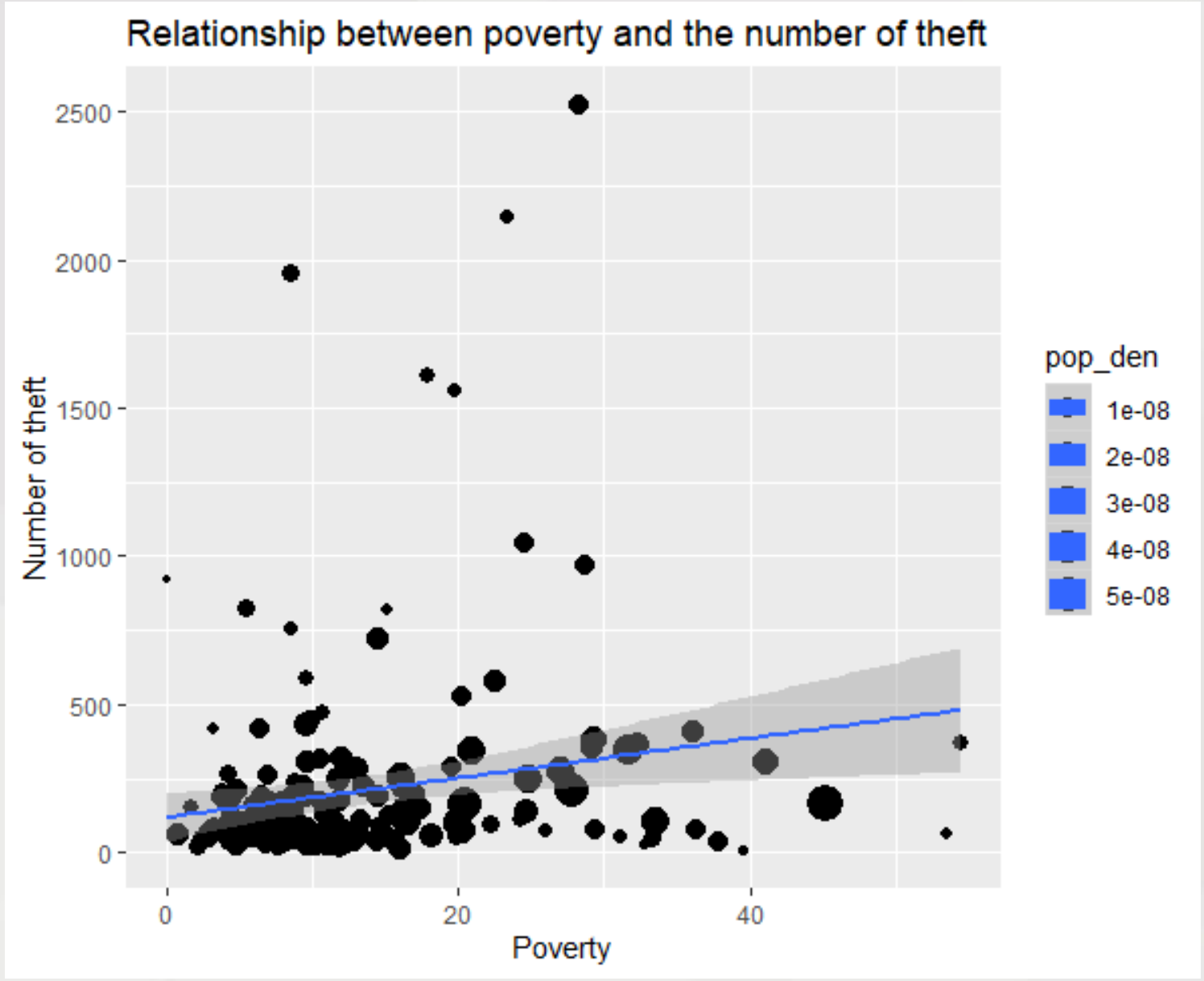
Poverty

1.> *TO PERFORM INNER JOIN*

```
dsPlot <- crime_SFO %>%
filter(Category=='LARCENY/THEFT')%>%
group_by(Tract)  %>%
summarise(total_crime=n()) %>%
inner_join(census_SFO, by = c("Tract" = "censustract"))
```

2.> *TO CREATE A SCATTER PLOT*

```
dsPlot %>%
ggplot(aes(x=p_poverty, y=total_crime, size=pop_den))+
geom_smooth(method = 'lm')+
geom_point()+labs(title="relationship with poverty",x="Poverty",y="Number of theft/larceny")
```



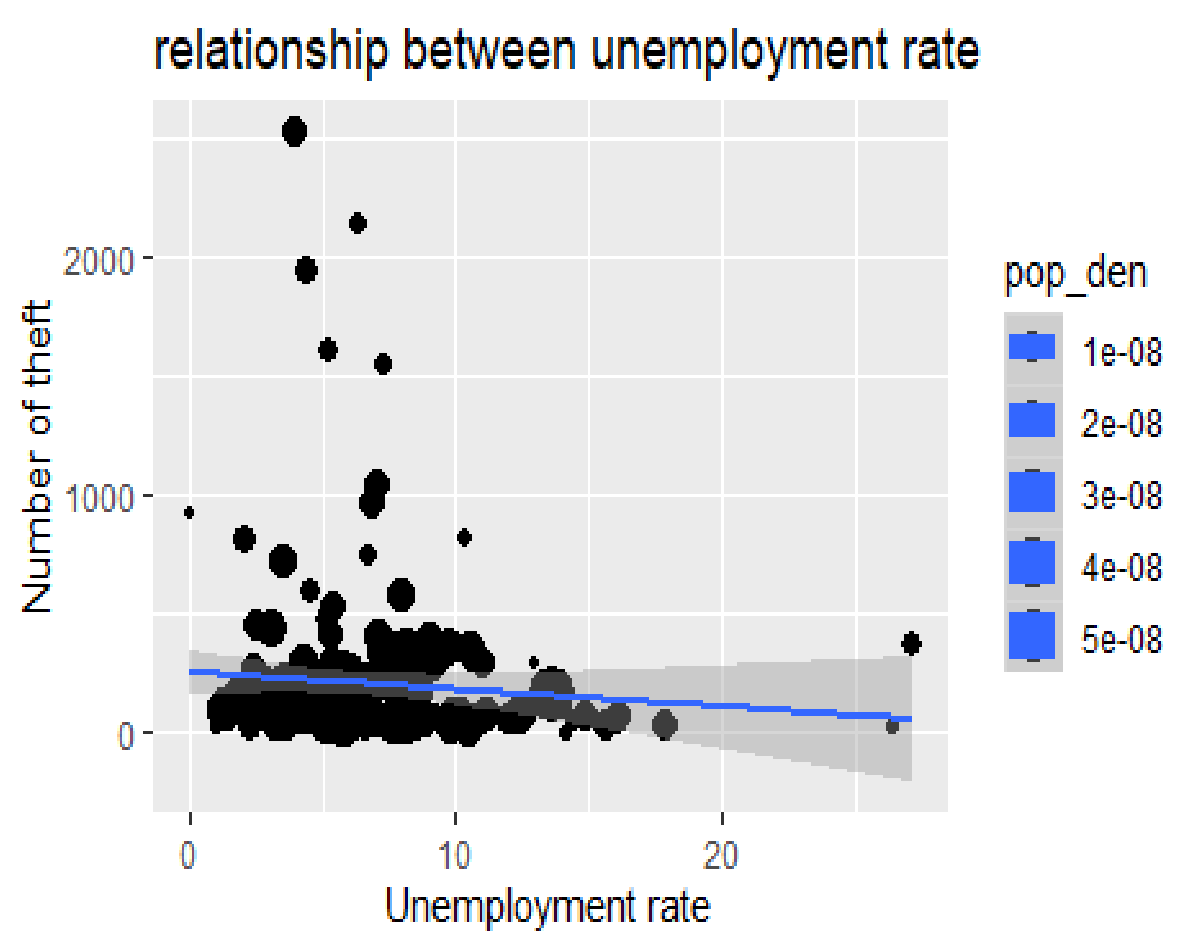
- Its is very evident that poverty does not stand behind to fuel the increase in crime rates.

- As the poverty increases, we see that the crime rates follow a positive growth.

Unemployment rate

```
Unemployment_rate<- crime_SFO%>%
  filter(Category=='LARCENY/THEFT')%>%
  group_by(Tract) %>%
  summarise(total_crime=n(),Unemployment=mean(Unemployment)) %>%
  inner_join(census_SFO, by = c("Tract" = "censustract"))%>%
  ggplot(aes(x=Unemployment, y=total_crime,size=pop_den))+
  geom_smooth(method = 'lm')+
  geom_point()+
  labs(title="relationship between unemployment rate and the number of crime",
x="Unemployment rate",
  y="Number of theft")

Unemployment_rate
```



- Common theory in public policymaking is that higher unemployment causes higher rates.
- Individuals who are underemployed are significantly less likely to be involved in a burglary than someone who is working full-time Underemployed people, however, are not much less likely to commit robbery.
- People who are unemployed and seeking work are no more likely to commit robbery or burglary than individuals with full-time jobs.

Age group

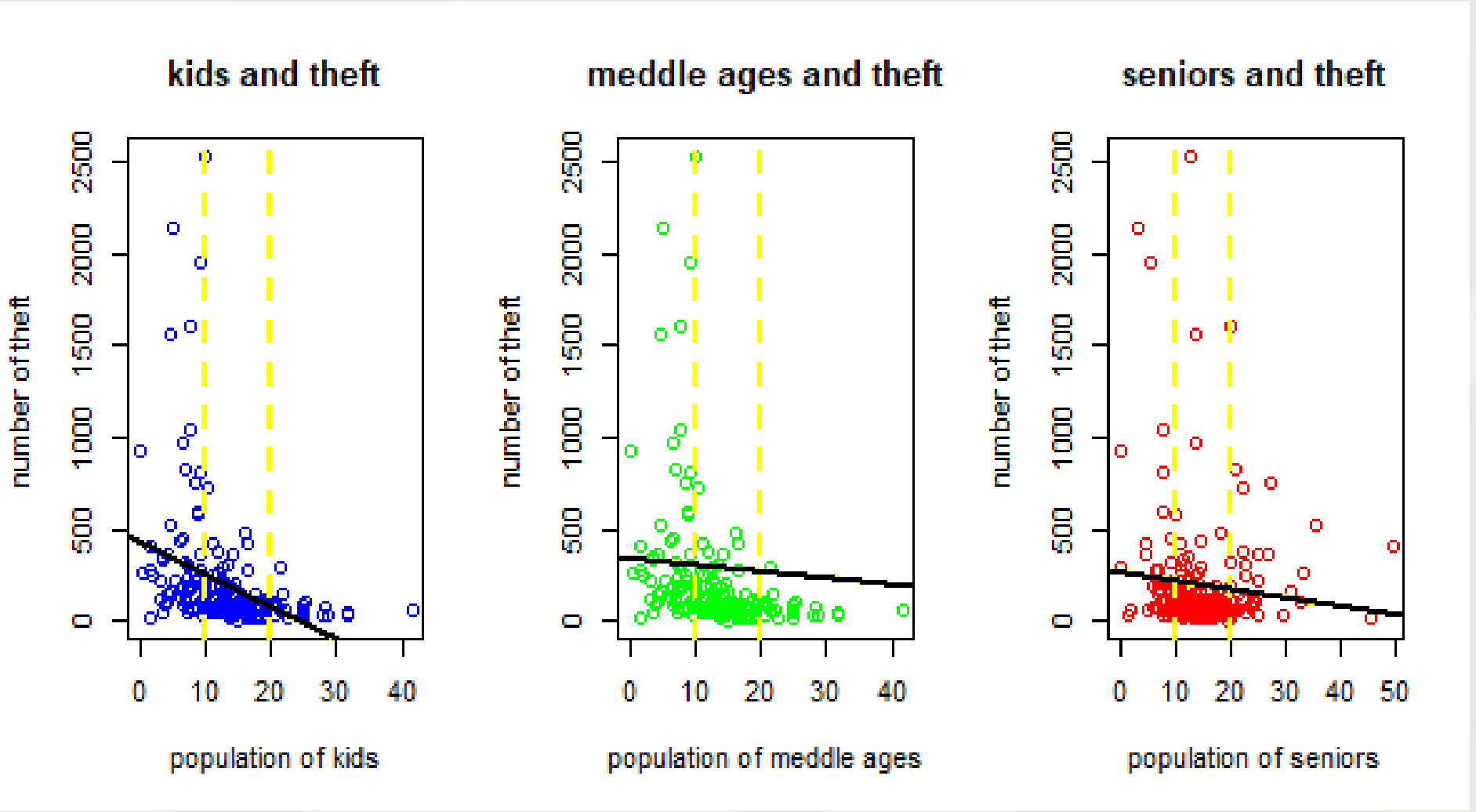
1.> *TO PERFORM INNER JOIN*

```
dsPlot <- crime_SFO %>%
filter(Category=='LARCENY/THEFT')%>%
group_by(Tract)  %>%
summarise(total_crime=n()) %>%
inner_join(census_SFO, by = c("Tract" = "censustract"))
```

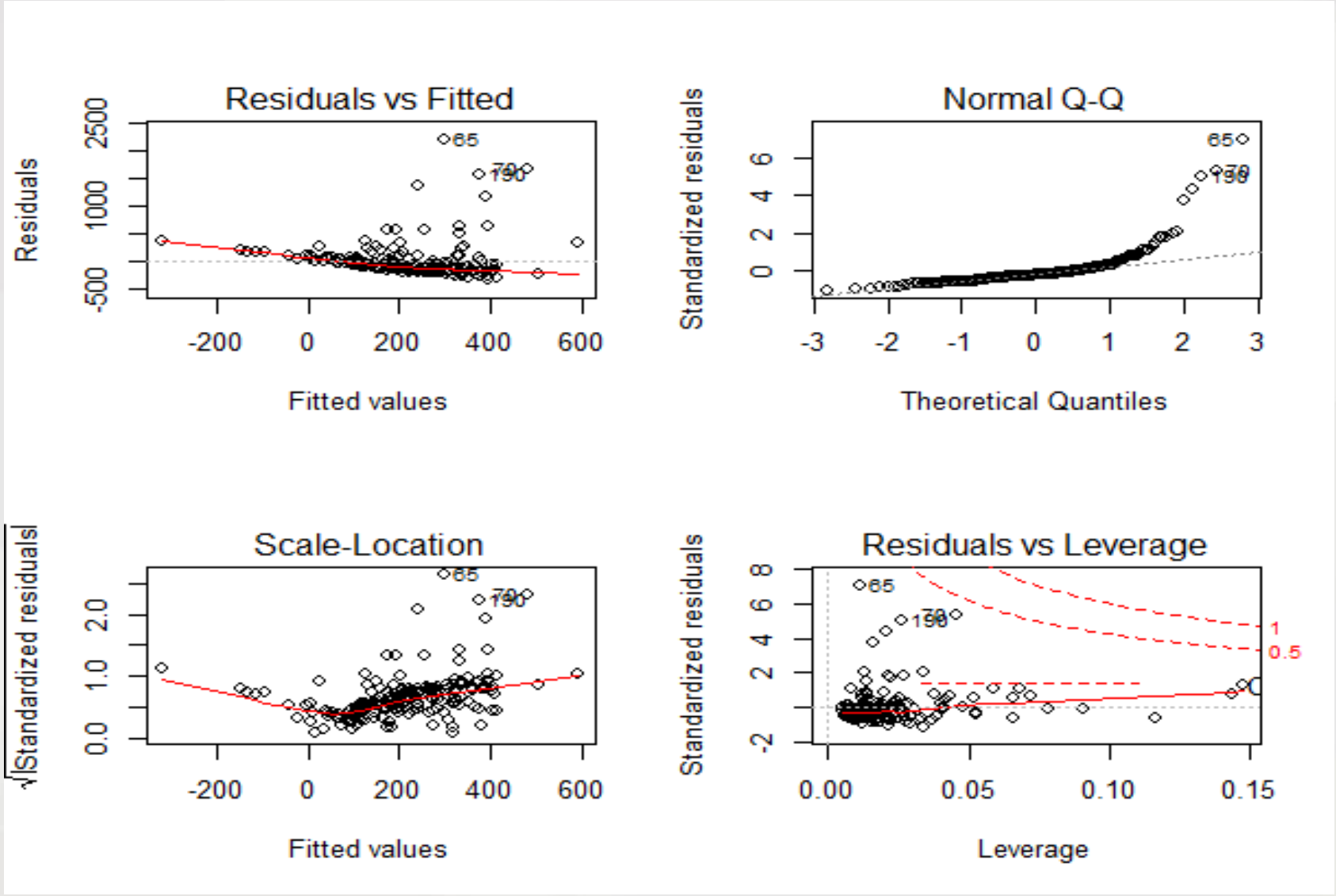
Age group

2.> *TO MERGE THREE SCATTER PLOT INTO ONE GRAPH*

```
  par(mfrow = c(1,3))
with(dsPlot, {
  plot(p_kids, total_crime, xlab="population of kids", ylab="number of theft", main = "kids and theft",col="blue")
  fit1<-lm(total_crime~p_kids)
  abline(fit1,v=10,lty=2,lwd=2,col="yellow")
  abline(fit1,v=20,lty=2,lwd=2,col="yellow")
  abline(fit1,lwd=2)
  plot(p_kids, total_crime, xlab="population of meddle ages", ylab="number of theft",main = "meddle ages and
theft",col="green")
  fit2<-lm(total_crime~med_age)
  abline(fit2,v=10,lty=2,lwd=2,col="yellow")
  abline(fit2,v=20,lty=2,lwd=2,col="yellow")
  fit2<-lm(total_crime~med_age)
  abline(fit2,lwd=2)
  plot(p_seniors, total_crime, xlab ="population of seniors", ylab="number of theft", main = "seniors and theft",col="red")
  fit3<-lm(total_crime~p_seniors)
  abline(fit3,v=10,lty=2,lwd=2,col="yellow")
  abline(fit3,v=20,lty=2,lwd=2,col="yellow")
  abline(fit3,lwd=2)})
```


kids and theft
number of theft
0
500
1000
1500
2000
2500
0
10
20
30
40
population of kids
meddle ages and theft
number of theft
0
500
1000
1500
2000
2500
0
10
20
30
40
population of meddle ages
seniors and theft
number of theft
0
500
1000
1500
2000
2500
0
10
20
30
40
50
population of seniors

Model testing



Time: days of week

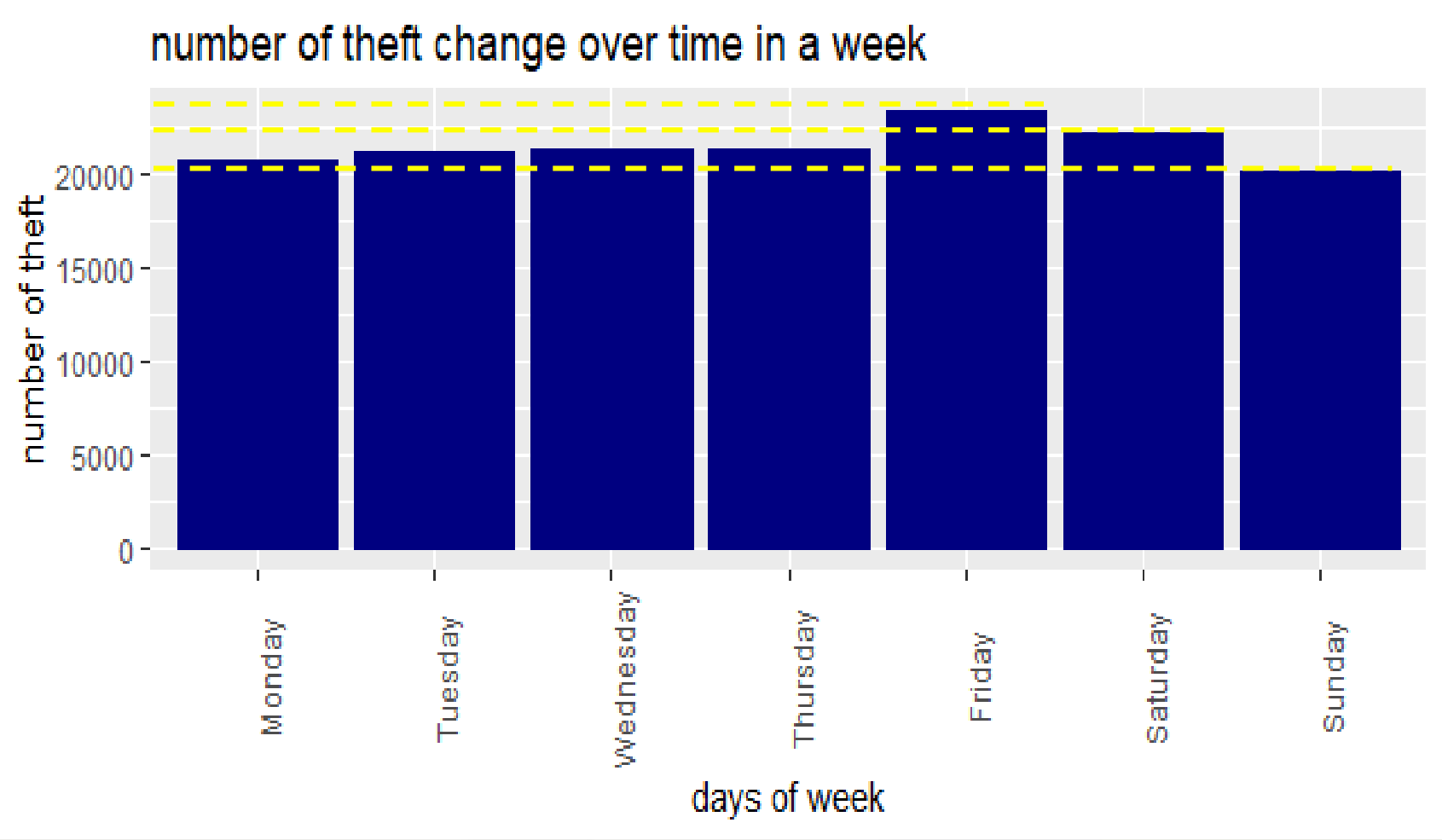
1.> *TO CALCULATE THE TOTAL NUMBER OF THEFT PER DAY*

```
dsPlot <- crime_SFO %>%
group_by(DayOfWeek)  %>%
summarise(total_crime=n()) %>%
mutate(DayOfWeek= factor(DayOfWeek, levels =
          c("Monday","Tuesday","Wednesday","Thursday","Friday","Saturday","Sunday")))
```

2.> *TO CREATE A SCATTER PLOT*

```
dsPlot%>%
ggplot(aes(x=DayOfWeek,y=total_crime)) +
geom_bar(stat='identity',color='navyblue',fill='navyblue')+
theme(axis.text.x = element_text(angle = 90))+
labs(title="number of theft change over time in a week", x="days of week", y="number of theft")
```

Based on days of week



Time: hour

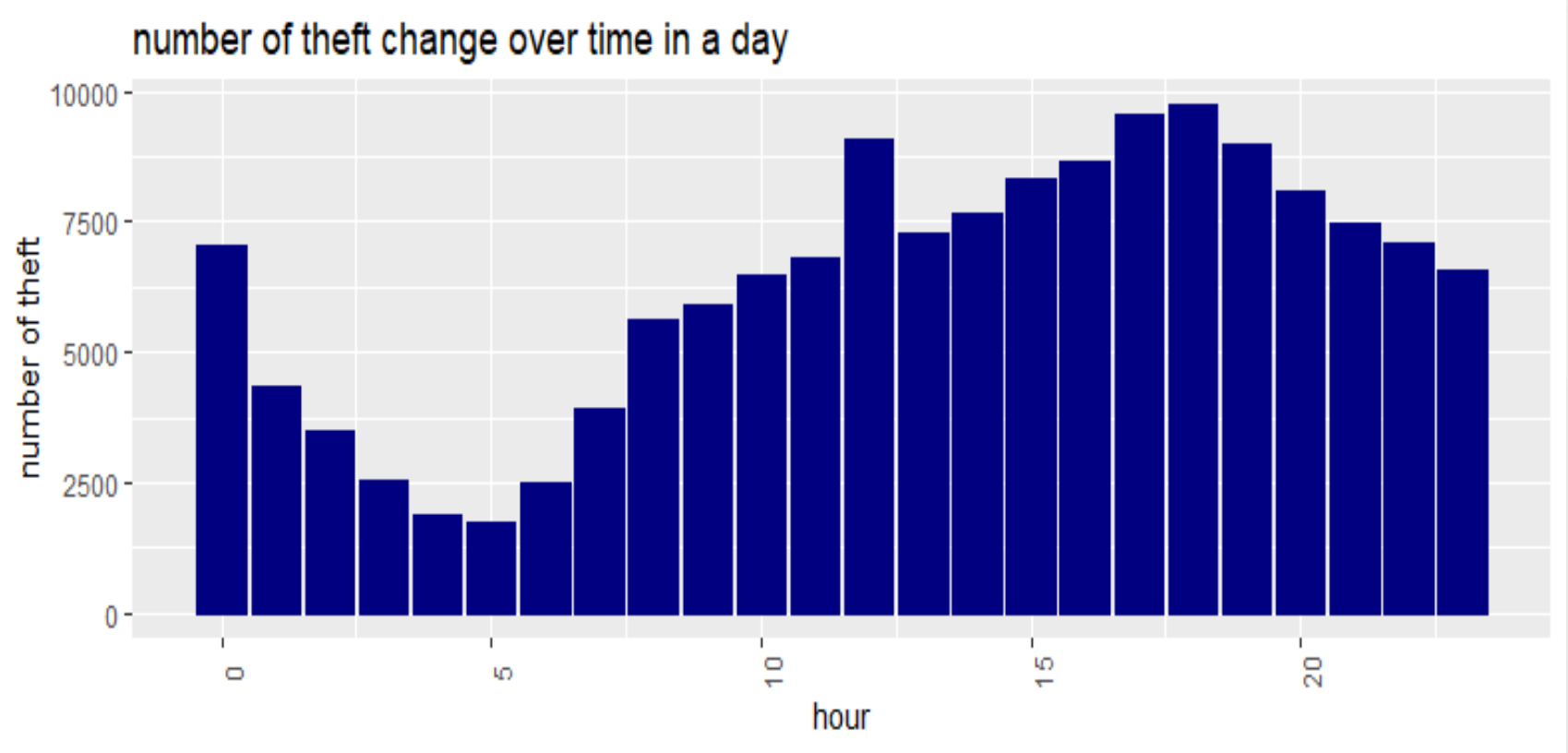
1.> *TO CALCULATE THE TOTAL NUMBER OF THEFT PER HOUR*

```
dsPlot2 <- crime_SFO %>%
group_by(Time)  %>%
summarise(total_crime=n())
```

2.> *TO CREATE A SCATTER PLOT*

```
dsPlot2%>%
ggplot(aes(x=Time,y=total_crime)) +
geom_bar(stat='identity',color='navyblue',fill='navyblue')+
theme(axis.text.x = element_text(angle = 90))+
labs(title="number of theft change over time in a day", x="hour", y="number of theft")
```

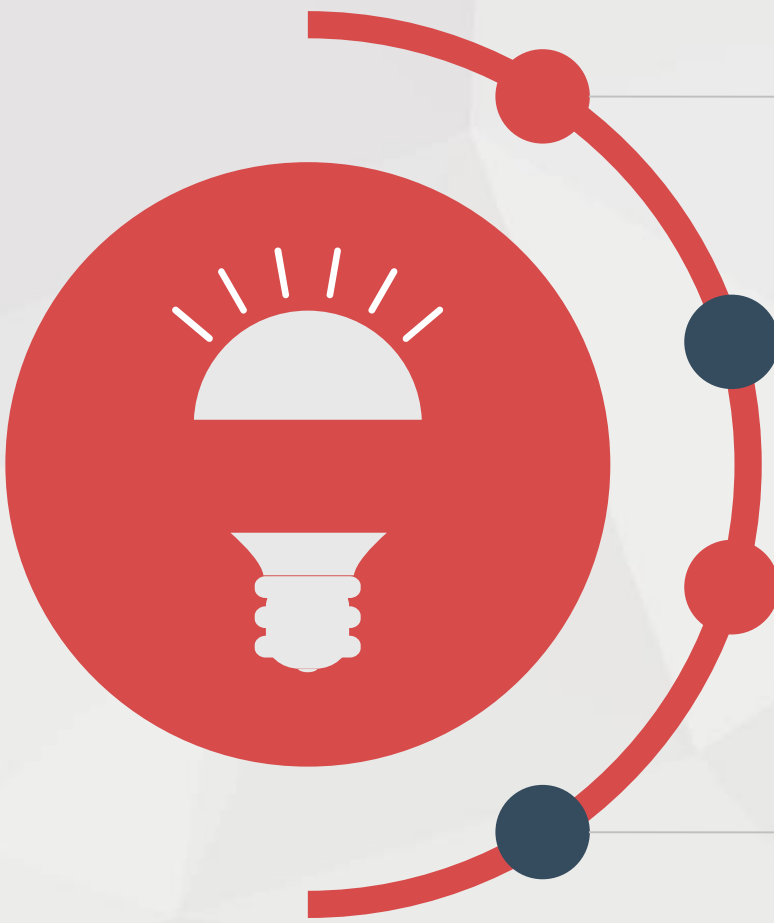
Based on hours



3

RECOMMENDATIONS

RECOMMENDATIONS



- Offer grant to poor people
- Improve kid's legal education
- Increase supervision on Friday
- Try to reduce unemployment rate

CONCLUSION

Thank you for your attention

Q & A